



# Chemical Identification with a Portable Raman Analyzer and Forensic Spectral Database\*

Nancy T. Kawai, InPhotonics, Inc., and James A. Janni, EIC Laboratories, Inc.

## Introduction

The last decade has brought much technical and commercial advancement in the field of Raman spectroscopy. More than ten years ago, the analytical community was immersed in the “Raman Renaissance” caused by the development and subsequent commercialization of FT-Raman spectroscopy. The coupling of interferometry with 1064 nm Nd:YAG excitation enabled new applications of Raman where previous attempts with visible excitation were hindered by enormous fluorescence backgrounds. It became apparent, however, that while very useful, FT-Raman was not the answer to all Raman applications due to the lower inherent sensitivity and the need for fairly large interferometer systems with precise moving mirrors.

Since the late 1980's, a more gradual but equally significant revival has been taking place in the realm of dispersive Raman spectroscopy. The development of highly sensitive CCD detectors, diode lasers, efficient notch filters, and compact spectrograph designs have catapulted dispersive Raman to the “What's New” headlines of analytical instrumentation columns. High performance systems have shrunk to bench-top designs, and price tags have dropped under \$100,000 and even to \$10,000. In only a few years, Raman has moved from the research laboratory to analytical environments and industrial process lines. Spectrometers are becoming more user-friendly with many of them packaged for dedicated applications.

One of these commercial systems is a portable Raman system designed primarily for on-site identification of unknown chemical species. The InPhotote™ is a portable Raman spectrometer carefully engineered to provide high optical throughput, excellent sensitivity, spectral range coverage of the critical “fingerprint” region, and ample resolution to differentiate crystalline solids. A focused fiber optic probe facilitates measurement of samples through glass and plastic packaging, eliminating the possibilities of contaminating precious forensic evidence or exposing operation personnel to potentially hazardous chemicals. There are many applications requiring portable instrumentation, including quality control measurements, environmental analysis, identification of hazardous materials, and forensic identification. The InPhotote™ was designed to collect high-quality Raman spectra in a compact, portable package, with flexible sampling and data analysis possibilities.

The circumstances of on-site analysis may require accurate identification within a few minutes even though the operators may have limited training to interpret spectral bands. Chemical identification can be facilitated by the availability of spectral reference databases or libraries for rapid comparison of data. Digital databases for many spectroscopic techniques are available in compilations of a few hundred to several thousand spectra. Unfortunately, owing to the relative newness of using Raman spectroscopy for analytical analysis, there are few spectral libraries available in digital format.

More so, those that are available tend to be instrument specific and may not produce as precise results as a truly standard library where spectra are corrected for instrument response and precise frequency calibration is maintained throughout the database. In collaboration with the Federal Bureau of Investigation,<sup>1</sup> a full spectral range and high resolution Raman library of forensic compounds has been developed. The goal of the library was to provide accurate identification in conjunction with any spectrometers' data regardless of optical design or detector, and to some degree, excitation wavelength. During the creation of the database, the standard searching algorithms used for infrared spectral searches were employed as they are widely available. Spectral searching of Raman data has not been widely used or studied, and the potential interferences in the search process merit further investigation. The reduction in identification precision and accuracy caused by fluorescence baselines, noise, and fixed spectral resolution were determined in the context of identifying solid and liquid samples. This note will discuss the compilation of the library, the outcome of the software investigations, how the optimal specifications were determined for the portable Raman spectrometer, and examples of their combined use in forensic applications.

\*An extended version of this application note appeared in *Spectroscopy* 15(10) 32-41 (2000).

<sup>1</sup>The Forensic Raman Spectral Library was constructed with the financial support of the FBI under contract JFBI98082.

*Table I. Comparison of Integrated Intensities of Cyclohexane Raman Peaks for Instrument Response Function Correction.*

Band Position (cm <sup>-1</sup> )	Spectrometer #1 Ratio	Spectrometer #2 Ratio	Reported Ratio (Ref. 1)
801	1.00	1.00	1.00
1028	0.62	0.57	0.65 ± 0.03
1267	0.47	0.47	0.52 ± 0.02
1444	0.55	0.57	0.65 ± 0.04
C-H stretches	7.72	7.44	7.38 ± 0.76

## The Forensic Raman Spectral Library

Raman data were collected on two RS2000 echelle spectrometers. The systems were equipped with stabilized 785 nm diode lasers as the excitation sources. Samples were measured with InPhotonics' focused fiber optic probes (RamanProbe™). Data were collected at 3.3 cm<sup>-1</sup> spectral resolution over a range of 200 - 3200 cm<sup>-1</sup> Stokes shift.

## Instrument Response Correction

From the onset of this project, the major goal was that the resulting database could be used with any properly calibrated spectrometer. The authors do not dispute that vendor-specific databases can still provide accurate identification of unknowns, however a "universal" database would generally provide more precise matching. The need to correct for instrument response is essential in Raman spectroscopy since the intensity axis is arbitrary. Factors that vary across the spectral region include detector response, spectrograph efficiency, and filter effects. In combination, these factors are unique to each individual instrument, and even distinctive between equivalent units from the same manufacturer. One can argue that data measured on a spectrometer today will be identical to the data measured on that spectrometer next year. This is not always true, however, as any change in a single component (e.g. after repair or replacement, or simply through aging) will affect the overall instrument response.

To remove instrument dependence, the total response curve must be determined. A NIST traceable standard irradiance source from Oriel Instruments (Stratford, CT) and the provided emission curve were employed. From the differences in the measured and actual irradiance curves, the total instrument response of the spectrometer was determined and used to mathematically adjust all the measured Raman data. Not surprisingly, changes in the collection geometry can dramatically affect the relative intensities in a Raman spectrum; careful attention was paid to recreating the Raman collection geometry in the white light calibration.

While accurate, the NIST traceable source is not very convenient to use on a regular basis. As such, many research groups have been investigating the use of secondary standards to determine instrument response.<sup>2</sup> Following the methods of Frost and McCreery,<sup>3</sup> the integrated relative intensities of cyclohexane Raman bands were measured immediately upon intensity calibration of the two different spectrometers. Although these ratios did not agree exactly with the reported values, the intensities were consistent with the errors associated with the calibration process. More importantly, the ratios could be used as a rapid check for subsequent drift in instrument response. Table I shows the cyclohexane band ratios of the two echelle spectrometers. For the subsequent data-collection phase of the project, the cyclohexane bands were measured and compared to these initial values; over the course of fifteen months of data

collection, the white light response correction held true.

## Data Collection and Treatment

Version 1.0 of the Forensic Raman Spectral Library contains spectra of 243 unique materials assembled in collaboration with forensic experts. Materials considered for inclusion fell into various categories:

- Explosives and decomposition products
- Over-the-counter and prescription drugs
- Illicit drugs and derivatives
- Household chemicals
- Solvents

Of the many candidates that were measured, less than ten materials exhibited strong fluorescence backgrounds that obscured the entire Raman spectrum when excited at 785 nm. These were excluded from the database. A reasonable number of materials fluoresced anywhere from slightly to moderately. The curved baselines and offsets were corrected using the interactive baseline correction function in GRAMS/32® from Galactic Industries, Inc. (Salem, NH). There was considerable deliberation prior to performing the baseline corrections as summarized in Table II.

In the end, it was concluded that background correction was warranted to

<sup>2</sup>Discussions at ASTM Raman Committee meetings, 1998-2000.

<sup>3</sup>K. Frost and R.L. McCreery, *Appl. Spectrosc.* **52** 1614-1619 (1998).

Table II. Potential Effects of Baseline Correction on the Reference Library.

Arguments Rejecting Baseline Correction	Arguments Supporting Baseline Correction
Fluorescence spectrum may be characteristic of the actual material.	Library search is more likely to match Raman bands rather than the fluorescence background.
In removing fluorescence, Raman features may be altered.	Dependence on excitation wavelength is somewhat removed.
Absence of fluorescence in the library match may confuse the operator.	Baseline correction improved the accuracy and speed of library matching.

meet the goals of the project and the database. For the majority of samples, exposure times were selected to achieve a signal-to-noise ratio (most intense peak:peak-to-peak baseline noise) of at least 10:1. Care was also exercised to ensure that the detector was not near saturation when measuring very strong scatterers. All data were checked and corrected for spurious spikes caused by the detection of cosmic rays and, occasionally, ambient room lights. Spectra of weak scatterers, where the signal-to-noise goals could not be met, were smoothed with a 3-term 9-point Savitzky-Golay function. The spectra that were previously corrected for strong fluorescence were particularly noisy, and these were subjected to an additional Fourier smoothing to further improve the signal-to-noise without distorting the spectral features. Figure 1 shows an example of the data treatment applied to one of the poorest quality raw spectra.

## Evaluating the Database

In a library search, a HQI (hit quality index) is calculated as the distance between the sample spectrum and each member of the library. Although there are different algorithms available in commercial software, the "correlation" search is the best general, full spectrum searching algorithm.<sup>4</sup> The HQI is calculated as the normalized least squares dot product between

<sup>4</sup>Spectral ID User's Guide, Galactic Industries, Inc., 77-80 (1999).

the unknown and the library spectrum after both spectra are centered about their respective mean. An HQI is computed for each library spectrum and ranked in increasing HQI value from best to worst match. The actual HQI values are only relevant within each search result; in the ideal case, the HQI of the top hit will be significantly lower than those of the remaining hits. It is important to remember that a library search will always provide a top hit even if the unknown material is not included in the library. Hence, the first 3 - 5 hits should be evaluated and not just the best match.

The "1st derivative correlation" algorithm is identical except that the differences are calculated on the first-derivative curves of the Raman spectra. This has the advantage that fluorescence backgrounds are removed from the distance calculation. An overview of the spectra in the database showed that many had offsets and curved baselines; these raised baselines would be exacerbated in real life samples due to the increased likelihood of impurities. The unpredictable shape of Raman baselines makes them difficult to correct with current automated software functions. Since correcting the baselines of analyte spectra should not be a requirement for rapid chemical identification, it was concluded that the "1st derivative correlation" algorithm was the best choice for the current application, and that the "correlation" algorithm could be used only for non-fluorescing samples or after an interactive baseline correction.

## Resolution

Unlike FT-IR instruments that can routinely provide spectral resolution better than the natural linewidths of solid and liquid phases, dispersive Raman spectrometers generally have a fixed resolu-

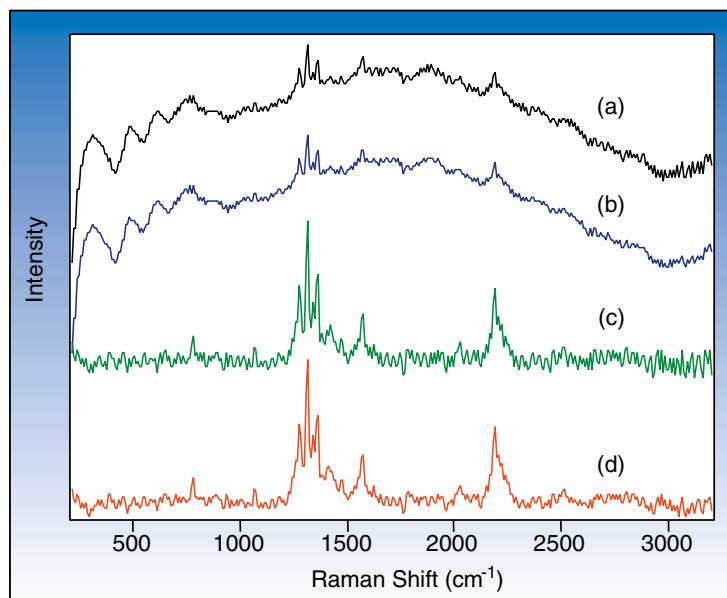


Figure 1. Data treatment applied to the spectrum of a weak scatterer exhibiting a fluorescence background. Shown are (a) a raw spectrum, and the spectrum after (b) a 3-term, 9-point Savitzky-Golay smoothing, (c) a multiple-point interactive baseline correction, and (d) an 80% Fourier smoothing.

tion for a given spectral region that is limited by the dispersion on the CCD array or by the linewidth of the excitation laser. Since higher resolution increases the complexity of instrumentation (including the size and addition of moving optics), determination of the minimum resolution for the portable Raman spectrometer was required.

To illustrate the importance of resolution in Raman spectral matches, a simple demonstration of the searching process was performed using synthetic spectra each containing a single band centered at 900, 902, and 904 cm<sup>-1</sup>. Three libraries were constructed to simulate databases measured at different spectral resolution; spectra at 8 cm<sup>-1</sup>, 16 cm<sup>-1</sup>, and 32 cm<sup>-1</sup> resolution were generated. The simulated spectra in Figure 2 visually show how more similar the three spectra appear as the measured resolution degrades. A subsequent search of the 900 cm<sup>-1</sup> band (at the corresponding resolution) resulted in the HQI values shown as part of the figure. It is apparent that, although each test experiment resulted in the proper first match, the HQI of the incorrect spectra increased enough at lower resolution to reduce confidence in the top match. In summary, the precision of the library search degraded as the measured resolution of the spectra decreased. This follows a recent infrared study that concluded that optimum library searches are performed when the measured spectral resolution is similar to the natural linewidths of the spectral features.<sup>5</sup> The natural linewidths of Raman bands are inherently narrower than those of infrared absorption bands, necessitating higher spectral resolution. For the 243 spectra included in the database, the natural linewidths ranged from 3 cm<sup>-1</sup> to 100 cm<sup>-1</sup>, with the typical linewidth of the crystalline powders being between 4-8 cm<sup>-1</sup>. This set the resolution goal for the portable Raman system.

### Frequency Calibration

The precision required to pinpoint a spectral feature is also related to the natural linewidth of that band. While the accuracy is determined by the frequency calibration of the spectrograph,

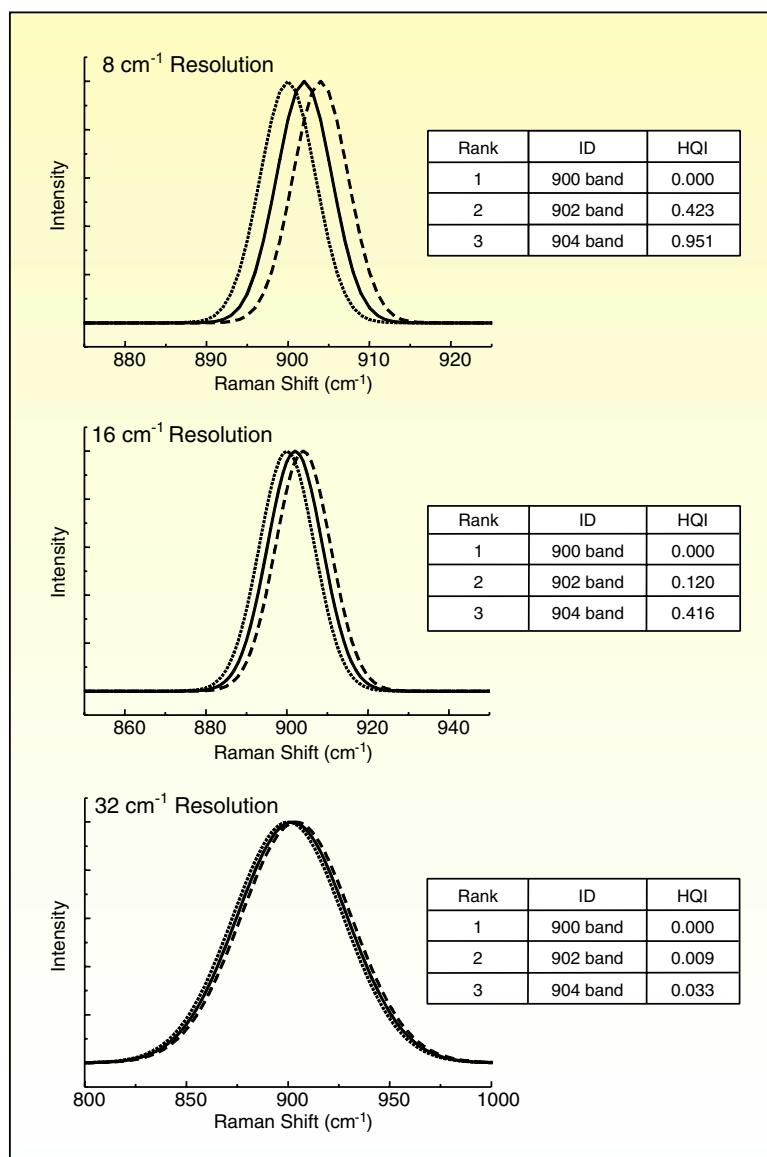


Figure 2. Spectral searching results for the 900 cm<sup>-1</sup> band when compared with a three-component library of synthetic spectra at resolutions of (a) 8, (b) 16, and (c) 32 cm<sup>-1</sup>.

the precision is determined by the size of the resolution elements (e.g. point spacing along the X-axis). Due to the unique design of the echelle spectrograph used to collect the library spectra, the resolution of the forensic database was 3.3 cm<sup>-1</sup>, but the precision was better than 0.4 cm<sup>-1</sup>. This very high precision is not typical of most dispersive Raman systems. The importance of frequency calibration and precision depends upon the number of spectral features for a given sample. Empirical testing of the library showed that, in cases where the resolution of the spectrometer was better than the natural linewidths of the materials under consideration, the frequency accuracy required was better than 3 cm<sup>-1</sup>. This limit was determined using the 1st

derivative correlation method; the methods not employing a first derivative calculation were more forgiving. This agrees with a reported study of the identification of pharmaceutical species by library searching methods.<sup>6</sup>

The identification of sugar is shown in Figure 3 to demonstrate this need for accurate system calibration; the 1st derivative correlation search fails completely with only a 3 cm<sup>-1</sup> miscalibration of the frequency axis. At lower

<sup>5</sup>A.S. Bangalor, J.C. Demirgian, A.S. Bopovai, and G.W. Small, *Appl. Spectrosc.* **53** 1382-1391 (1999).

<sup>6</sup>R.L. McCreery, A.J. Horn, J. Spencer, and E. Jefferson, *J. Pharm. Sci.* **87**(1) 1-8 (1998).

spectral resolution, where the line-widths are now determined by instrument rather than sample, the resolution itself becomes the more important factor toward misidentification rather than frequency precision and accuracy.

To achieve accurate and precise identification of solid samples, and crystalline species in particular, frequency calibration within  $3\text{ cm}^{-1}$  is essential.

### Baseline and Noise Effects

In general, increased noise does not prevent accurate identification provided that the Raman spectral features are clearly defined over the baseline noise. Higher noise will reduce the precision of the spectral search; the HQI's of the true match will not be as clearly separated from the subsequent hits. For searches that do not perform a first derivative calculation, the overall intensity contributions from the baseline noise caused some confusion with other noisy spectra in the database, hence the decision to smooth the very noisy spectra prior to their inclusion in the library.

Baseline curvature and offset are generally caused by the sample itself and little can be designed in an instrument to avoid these interferences. Testing confirmed that performing a search using an algorithm that does not calculate a first derivative causes the

software to match the baseline offset rather than the narrower Raman bands and justifies baseline correction of the library spectra.

### Designing the Portable Raman Spectrometer

**A**ccording to the tests conducted with the Forensic Raman Spectral Library, the resolution goal was set for the portable system as, ideally,  $4 - 6\text{ cm}^{-1}$  for highest precision and accuracy for the identification of crystalline solids. The question of spectral range coverage was the next specification under consideration. Increased spectral range is intrinsically tied to resolution limits on a linear spectrograph. Since it was determined that relatively high resolution was required for this application, the range needed to be truncated from that of the echelle spectrograph. Although the higher frequency range clearly contains information about C-H, O-H, and N-H functionalities, over 95% of the bands lie between 300 and  $1700\text{ cm}^{-1}$  and could therefore be identified by bands in this region alone. The only exception to this generality occurs for simple cyanide salts, where the  $\text{C}\equiv\text{N}$  stretch occurs at about  $2230\text{ cm}^{-1}$ . Over a 1024 pixel array, the expansion of the spectral range to include the cyanide region would seriously compromise the overall spectral resolution. Since the major goal of the instrument was to provide

high identification accuracy on solids and liquids, it was decided that the standard spectral range could be limited to below  $1800\text{ cm}^{-1}$  with the resolution as high as possible given the set detector array size.

All other specifications were determined in compromise with cost, size and weight. The overall package is portable by a single user ( $16'' \times 10'' \times 9''$ , 20 lbs.). The system was designed for operation with 110V AC power or by 12V DC power. The latter enables the spectrometer to be operated from a vehicle lighter jack or a portable battery pack. The optical components provided spectral range coverage of  $1600\text{ cm}^{-1}$  with a resolution of  $6\text{ cm}^{-1}$ . The 1024 x 128 pixel CCD array was TE-cooled to  $-25^\circ\text{C}$  for excellent sensitivity. A frequency-stabilized 785 nm diode laser was incorporated into the package to provide a narrow-line excitation source that would not diminish the spectral resolution.

### Testing the Complete Solution

**T**he InPhotote™ and Raman Forensic Spectral Library were subjected to a variety of samples to demonstrate their ability to identify unknown species in both solid and liquid phases. These search results are shown in Figure 4. A spectrum of epinephrine was measured for 50 seconds

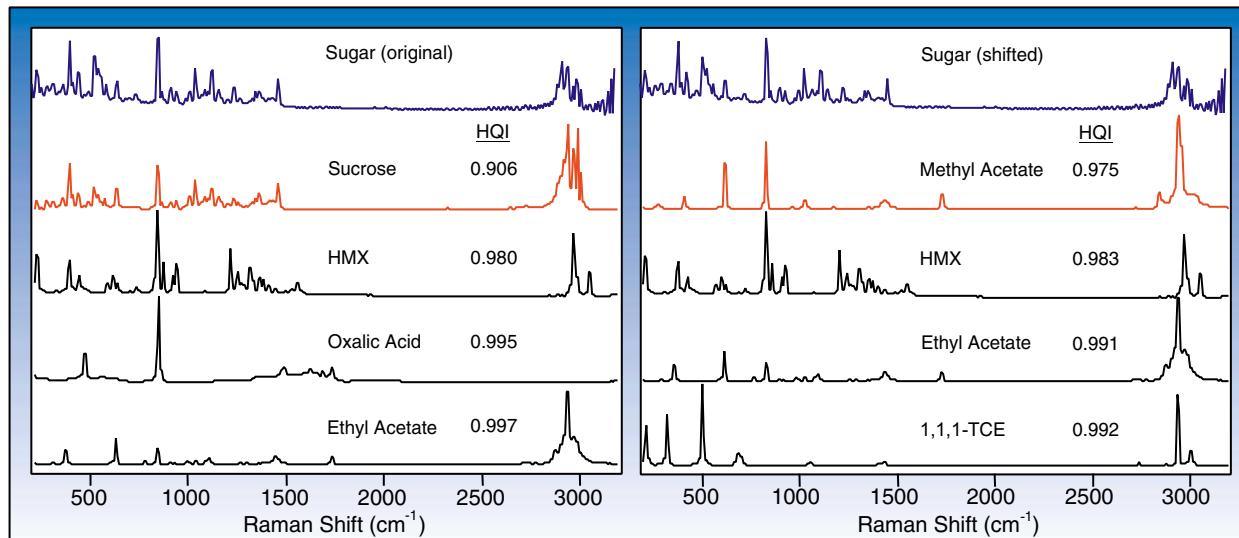
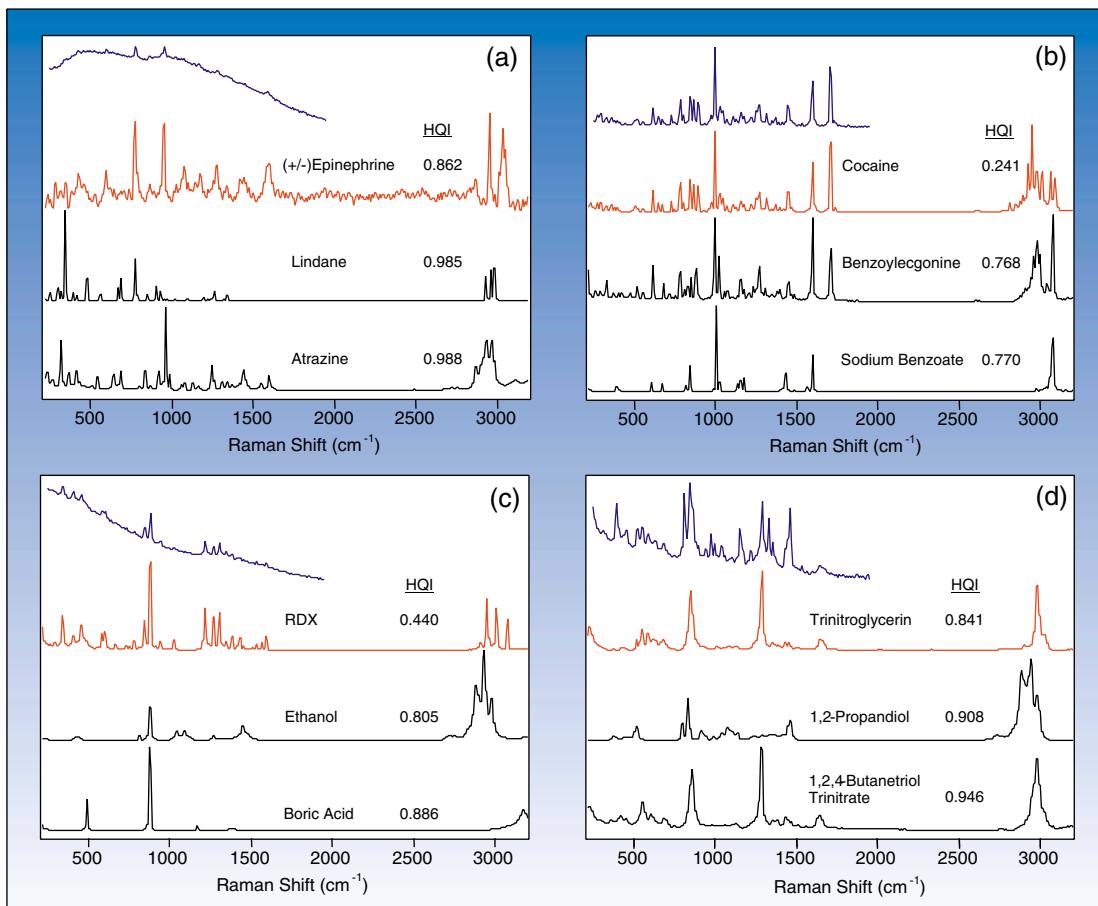


Figure 3. Library search results for (left) a sugar spectrum, and (right) a sugar spectrum after a  $3\text{ cm}^{-1}$  shift along the frequency axis. The correct match appeared 9<sup>th</sup> on the hit list for the shifted spectrum.



*Figure 4. Identification of narcotics and explosives using the portable Raman spectrometer and forensic spectral library. Shown are analyses of (a) epinephrine (50-sec data collection), (b) cocaine (10-sec), (c) RDX (5-sec), and (d) nitroglycerin (20-sec through a plastic vial). All searches were conducted with the first-derivative correlation algorithm without any postcollection data treatment.*

and subjected to identification through the Spectral ID™ software (Version 2.01, also from Galactic Industries, Inc.). The highly fluorescent, weak scattering sample was identified without difficulty using the 1st derivative correlation algorithm. Having the operator interactively correct the baseline enabled the use of the correlation algorithm, which resulted in better separation of the top hit from the remaining matches. A ten second spectrum of cocaine was also correctly differentiated from several derivatives that were included in the library.

A blind sample from an explosives laboratory was measured for five seconds. The results of the library search clearly identify the sample as RDX. Another blind sample was measured through a plastic beaker by holding the focused probe against the container. Even with the interfering polymer bands, the sample was correctly identified as a nitroglycerine explosive.

## Conclusion

A well-controlled library of Raman spectra was developed for the identification of forensic chemicals. Measured at better than  $4\text{ cm}^{-1}$  resolution from  $220 - 3200\text{ cm}^{-1}$ , the library was designed for use with spectra collected on any properly-calibrated Raman spectrometer system. The information gained during the construction and testing of the library provided valuable insight into the minimum requirements for a new spectrometer. Consideration of the most important specifications, with some compromises for cost and size along the way, resulted in the development of the InPhotote™, a portable Raman spectrometer designed for field identification of unknown chemicals.

If you would like additional copies of this note, or if you would like to speak with a member of our technical staff, please contact us at:



111 Downey Street  
Norwood, MA 02062  
Tel. (781)440-0202  
Fax (781)551-0283

[www.inphotonics.com](http://www.inphotonics.com)  
[info@inphotonics.com](mailto:info@inphotonics.com)